Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA
Fakultas MIPA, Universitas Negeri Yogyakarta, 16 Mei 2009

DISTRIBUSI PELUANG BANYAK KOMPONEN DALAM *RANDOM FUNCTION*

S. Padmadisastra

Jur. Statistika, Fmipa – Universitas Padjadjaran
Bandung
(s_padmadisastra@yahoo.com)

Abstrak

Dalam makalah ini dibahas masalah yang terkait dengan *random functions*; yaitu mengenai banyak komponen yang terbentuk. Solusi aproksimatif untuk masalah ini diperoleh melalui pendekatan dalam *population genetics*.

PENDAHULUAN

Perhatikan sebuah set bilangan bulat, $S_n = \{1,2,\cdots,n\}$. Misalkan $f(x)$ merupakan sebuah fungsi pada *S into S*. Sebuah *random function* menetapkan, secara independen, kepada tiap bilangan $i \in S_n$ sebuah *image* $j \in S_n$, dengan peluang sama; yaitu sebesar $1/n$. Tentunya ada sebanyak n^n fungsi seperti ini. Dalam hal *one* – to – *one* *mapping* , setiap fungsi merupakan sebuah permutasi dari bilanga – bilangan dalam S_n. Sembarang *mapping function* f menghasilkan sebuah dekomposisi dari set S_n kedalam beberapa subset *disjoint*, *nonnull*, *minimal* yang tertutup dibawah *mapping*, jadi

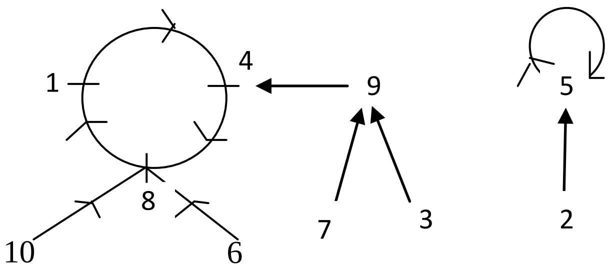
$$S_n = \omega_1 + \omega_2 + \cdots + \omega_k$$

dengan

$$f(\omega_i) \subset \omega_i \text{, dan } f^{-1}(\omega_i) \subset \omega_i .$$

Subset-subset yang terbentuk disebut sebagai “tree” atau “components of the mapping”. Fungsi f seperti ini dikatakan mendekomposisi set S_n menjadi k buah komponen. Dalam hal permutasi, maka komponen-komponen berupa *cycles* dari permutasi. Sebagai contoh, misalkan $S = \{1,2,3,4,5,6,7,8,910\}$, dan $f(1) = 4$, $f(2) = 5$, $f(3) = 9$, $f(4) = 8$, $f(5) = 5$, $f(6) = 8$, $f(7) = 9$, $f(8) =1$, $f(9) = 4$, $f(10) = 8$; dalam hal ini banyak komponen adalah $k = 2$ dan $\omega_1 = \{1,3,4,6,7,8,9,10\}$, $\omega_2 = \{2,5\}$.

Sebuah *random function* dapat dinyatakan sebagai sebuah *directed graph*, sebuah “tree rooted in a cycle”. Berikut diperlihatkan sebuah *random function* secara grafis:



Gambar 1. Penyajian grafis sebuah *random function*

Beberapa besaran yang terkait dengan *random function* telah banyak diungkap. Metropolis dan Ulam (1953) mempermasalahkan rerata banyak komponen (E). Kruskal (1954) menjawabnya dengan sebuah solusi berikut:

$$E = \sum_{m=1}^{n} \frac{n!}{(n-m)!mn^m}$$

Lebih lanjut Kruskal menyatakan bahwa apabila *n* bertambah besar, maka *E* menjadi

$$E = \frac{1}{2}\left(\log 2n + C\right) + o(1)$$

Dengan *C* merupakan sebuah konstanta Euler; *C* = 0.5772…. Katz (1955) menyatakan peluang bahwa fungsi *indecompasable* (komponen tunggal) adalah

$$P(k=1) = \left((n-1)!/n^n\right)\sum_{m-0}^{n-1} n^m / m!,$$

dan untuk $n \to \infty$ menjadi

$$P(k=1) \cong \left(\frac{\pi}{2n}\right)^{1/2}.$$

Harris (1960), dan Kolchin (1976) memperlihatkan bahwa distribusi peluang dari vector dengan elemen ke-*i*, k_i, menyatakan banyak komponen berukuran *i*; $(k_1, k_2, \cdots, k_n)$, adalah

$$P(k_1, k_2, \cdots, k_n) = \frac{n!e^n}{n^n} \prod_{j=1}^{n} \left\{ \frac{p(j)^{k_j}}{j^{k_j} k_j!} \right\}, \qquad \ldots (1)$$

dengan

$$p(j) = e^{-j} \sum j^i \Big/ i!$$

Distribusi banyaknya komponen, sebutlah *k*, $k = \sum k_i$, dapat diperoleh secara prinsip dari distribusi vector di atas, tetapi ternyata tidak sederhana. Sebuah formula rekursif, yang memungkinkan perhitungan dilakukan, dikemukakan oleh Ross (1981). Sebuah solusi approksimatif, diperoleh dari *Population Genetics*, diusulkan sebagai berikut.

Deskripsi mengenai sebuah sampel *gene* dari sejumlah individu haploid, dalam *Population Genetics*, di perlihatkan melalui *frequency spectrum* dari *allele*; yaitu berapa banyak *allele* yang diwakili oleh sejumlah *gene* tertentu, dalam sampel tersebut. Sebagai sebuah illustrasi, misalkan sebuah sampel terdiri dari lima buah *gene* berikut $A_1 A_2 A_3 A_1 A_2$, maka frekuensi *allele* yang diwakili dua buah *gene* ada dua *allele*; yaitu A_1 dan A_2, dan hanya satu yang diwakili satu buah *gene*, A_3

Ewens (1972) menyatakan bahwa distribusi peluang *frequency spectrum* sebuah sampel yang terdiri dari *n* individu adalah:

$$P_n(k_1, k_2, \cdots, k_n) = \frac{n!}{\theta_{(n)}} \prod_j \left[\theta^{k_j} \Big/ \left(j^{k_j} k_j!\right)\right], \qquad \ldots (2)$$

Dan peluang banyak *allele*, $k = \sum k_i$, adalah

$$P(k=i) = \theta^i \mid S_n^i \mid / \theta_{(n)}.$$

Dalam formula di atas S_n^i menyatakan jenis bilangan Stirling pertama, dan *ascending factorial* $\theta_{(n)} = \theta(\theta+1)\cdots(\theta+n-1)$.

Formula (2), dikenal sebagai Ewens Sampling Formula (ESF), berlaku bagi sebuah *random permutation* dengan komponen berupa *cycle – cycle* dalam permutasi, dan dalam hal ini θ = 1 sesuai dengan hasil – hasil dari *random permutation*. Sedangkan untuk *random functions*, formula (2) tidak berlaku eksak hanya berlaku secara asymptotis saja. Lebih lanjut, Aldous (1985) menjelaskan bahwa keberlakuan asymptotis (2) terhadap (1) adalah untuk θ = ½.

PEMBAHASAN

Dalam tabel berikut diperlihatkan kesesuaian hasil perhitungan (1) secara simulasi dengan nilai eksak, melalui cara rekursif Ross, dan nilai asymptotiknya untuk berbagai nilai *n*, yaitu 50, 100, 150, dan 200.000, dan beberapa nilai *k*. Simulasi dikerjakan sebanyak 20.000 kali.

Tabel 1 (a) simulasi, (b) eksak, dan (c) approksimasi distribusi *K* untuk beragam *n*

n		*k*				
		1	2	3	4	5
50	*a*	0.1713	0.3270	0.2815	0.1488	0.0534
	b	0.1709	0.3262	0.2817	0.1485	0.0543
	c	0.1256	0.2814	0.2896	0.1844	0.0823
100	*a*	0.1228	0.2747	0.2878	0.1864	0.0872
	b	0.1221	0.2756	0.2886	0.1889	0.0875
	c	0.0887	0.2297	0.2792	0.2141	0.1171
150	*a*	0.1031	0.2355	0.2886	0.2058	0.1069
	b	0.1002	0.2466	0.2847	0.2073	0.1078
	c	0.0724	0.2022	0.2675	0.2250	0.1360
200	*a*	0.0862	0.2252	0.2824	0.2132	0.1202
	b	0.0870	0.2267	0.2789	0.2177	0.1220
	c	0.0627	0.1841	0.2575	0.2300	0.1483

n		*k*				
		6	7	8	9	10
50	*a*	0.0141	0.0033	0.0005	0.0001	0.0000
	b	0.0148	0.0031	0.0005	0.0001	0.0000
	c	0.0276	0.0073	0.0016	0.0003	0.0000
100	*a*	0.0304	0.0084	0.0019	0.0004	0.0001
	b	0.0308	0.0086	0.0020	0.0004	0.0001
	c	0.0490	0.0164	0.0045	0.0010	0.0002
150	*a*	0.0426	0.0132	0.0035	0.0009	0.0001
	b	0.0429	0.0137	0.0036	0.0008	0.0002
	c	0.0633	0.0237	0.0074	0.0019	0.0004
200	*a*	0.0500	0.0168	0.0052	0.0011	0.0001
	b	0.0526	0.0182	0.0052	0.0013	0.0003
	c	0.0739	0.0297	0.0099	0.0028	0.0007

Tabel di atas memperlihatkan bahwa hasil simulasi dengan perhitungan eksak sangat bersesuaian., untuk berbagai *n* dan *k*. Sampai dengan digit pertama, perbedaan antara hasil perhitungan aproksimasi dengan eksak atau simulasi, untuk *n* besar dan *k* besar Nampak tidak terlalu jauh berbeda. Hal lain yang diperlihatkan dalam table adalah bahwa secara umum, baik hasil simulasi, eksak, maupun pendekatan memperlihatkan gejala sama; yaitu makin besar *k* makin kecil peluang.

PENUTUP

Telaah lebih lanjut diperlukan untuk mendapatkan suatu bentuk analitis mengenai batas atas atau bawah dari perbedaan antara hasil perhitungan eksak dengan aproksimasi. Selain itu, beberapa masalah lain, seperti rata-rata banyak komponen yang terbentuk menarik untuk ditelaah lebih lanjut.

DAFTAR PUSTAKA

Aldous, D. J. 1985. Exchangeability and related topics, in "*Lecture Notes in Mathematics*" (A. Dold and B Eckmann, eds.) 1117, 71 – 198. Springer – Verlag, Berlin.

Ewens, W.J. 1972. The sampling theory of selectively neutral alleles, *Theor*. *Pop*. *Biol*. 3, 87 – 112

Harris, B. 1960. Probability distribution related to random mappings, *Ann*. *Math*. *Stat*. 31, 1045 – 1062

Katz, L. 1955. Probability of indecomposability of a random mapping function, *Ann*. *Math*. *Stat*. 26, 512 – 517.

Kolchin, V. F. 1976. A Problem of the allocation of particles into cells and random mappings, *Theor*. *Prob*. *Appl*. 21, 48 – 63.

Metropolis, N., and Ulam, S. 1954. A property of randomness of an arithmetical function. *Amer*. *Math*. *Monthly*, 61, 252.

Kruskal, M. D. 1954. The expected number of components under a random mapping function, *Amer*. *Math*. *Monthly*, 61, 392 – 397.

Ross, S.M. 1981. A Random Graph, *J*. *Appl*. *Prob*. 18, 309 – 315.